Real-time social distance measurement and face mask detection in public transportation systems during the COVID-19 pandemic and post-pandemic Era: Theoretical approach and case study in Italy

Marco Guerrieri [a,*], Giuseppe Parla [b]

[a] *DICAM (Department of Civil, Environmental and Mechanical Engineering), University of Trento, Via Mesiano 77, 38123 Trento, Italy*
[b] *ISMET (Mediterranean Institute for Transplantation and Advanced Specialized Therapies), via Tricomi 5 90127, Palermo, Italy*

ARTICLE INFO

ABSTRACT

Due to its remarkable learning ability and benefits in several areas of real-life, deep learning-based applications have recovered to be a research topic of great importance in the last few years. This article presents a method devoted to guaranteeing safety conditions in public transportation systems (PTS) during the COVID-19 pandemic and post-pandemic era. The paper describes a viable real-time model based on deep learning for monitoring social distance between users and detecting face masks in stop areas and inside vehicles of public transportation systems. Detections are made using the deep learning approach and YOLOv3 algorithm. The safety rule violations are represented by red bounding boxes and red circles in a bird's eye view as output of the video surveillance analysis. The datasets used to train the neural network are the "Caltech Pedestrian Dataset" and the "COVID-19 Medical Face Mask Detection Dataset". Metrics, such Loss Accuracy, and Precision, obtained in the testing process of the neural network were used to evaluate the performance of the model in detecting users and face masks. The proposed method was recently tested in the Public Transportation System of the Municipality of Piazza Armerina (Italy). The results show a significant reliability of the method in detecting real-time interactions between users of the PTS in terms of over-time variations in their mutual distancing, as well as in recognising cases of violation of the imposed social distancing and FFP2 face mask use.

Introduction

The World Health Organization (WHO) on 11 March 2020 claimed that the international outbreak of the new coronavirus SARS-CoV-2 infection was to be considered a pandemic. Coronavirus disease (COVID-19) can cause multiple symptoms with different levels of severity, ranging from asymptomatic disease to acute respiratory failure, even death. The coronavirus can be spread through the air by droplets (5 to 10 μm) and fine aerosols (smaller than 5 μm) exhaled from infected individuals when breathing, speaking, coughing, or sneezing (Prather et al. 2020). The virus is highly contagious, the incubation period varies from a minimum of two days to a maximum of fourteen days depending on the variant considered; the median incubation period was estimated to be 5.1 days (Lauer et al., 2019). The disease spread quickly around the world and globally had a shocking impact on the population's health, welfare and economics. In accordance with the warnings of the WHO, physical distancing between people helps limiting COVID-19 infections from spreading widely and thus it has been taken as a key strategy for reducing the number of infected people since the beginning of the pandemic. Following the advice of the World Health Organization, the public authorities have proposed 1 to 2 m of interpersonal distancing. Another effective way to prevent COVID-19 infections is to wear face masks (Tang et al., 2009) and therefore, in several countries, wearing face masks has become compulsory in crowded areas (e.g. in public transportation systems). As the spread of the virus mainly depends on the interaction among people, social distancing and wearing face masks can reduce the spread of the disease. Reducing the rate of spread of the disease (e.g. in terms of the effective reproductive number R0) means less infection, a smaller number of deaths and even less undesirable effects on other domains of life. Therefore, health authorities around the world have suggested wearing face masks, maintaining adequate "social distancing" in indoor environments, (at least one meter) and avoiding crowded places even outdoors. These are the only non-pharmaceutical public health measures that, if effectively respected by the majority of

---

the population, provide indirect benefits like a lower probability of oversaturation of the health system of a given country and a significant reduction in the rate of infections and deaths (Fig. 1). The evolution of the spread of an infectious disease can be analysed through epidemiological models, the best known of which is the SIR model by Kermack and McKen-drick (Kermack & McKendrick, 1927). Recently, Eksin et al. (2019) have developed an epidemiological model largely based on the SIR model but suitably modified to include the effect that compliance with an appropriate social distancing can lead to the evolution of the spread of the infection in a given examined population. Hence, with the use of adequate epidemiological models, it is possible to formulate a reliable estimate of the future evolution of the pandemic and then adequately program the types and intensity of health measures to contrast the phenomenon.

Public transportation systems (buses, trains, trams, trolleybuses, ferries, rapid transit, etc.) play a crucial role in the transmission of the virus (De Vos, 2020), given the densely crowded conditions normally found at station stops and even inside vehicles (Grant and Booth, 2009; Corriere, et al. 2013; Guerrieri, 2019; Calderón Peralvo et al., 2022). Therefore, in public transportation systems, it may be useful to introduce suitable techniques for monitoring the social distance between users and detecting whether users wear face masks correctly along the journey (Park and Kim, 2021). Clearly enough, conventional monitoring techniques cannot be used for this type of control; instead, smart control measures can play a fundamental role in this field. Today artificial intelligence (AI) is applied to many fields of life and numerous are the applications in the sector of transportation engineering. The ability to "learn" useful information, to propose decision-making processes and to produce useful results are the most surprising characteristics of deep learning. These distinctive features of deep learning will lead to intelligent automation of transportation systems and reduction of operating costs. Computer vision, moreover, allows the computer to evaluate visual data. The recent evolution of Deep Learning (DL) algorithms makes it easy for a computer to analyse digital data and offers the possibility of detecting a huge number of objects of interest in an increasingly reliable way starting from the analysis of images or videos. Consequently, the DL can be used as a method for detecting whether users wear face masks correctly, for analysing the instantaneous distancing between users of local public transport systems and identifying cases in which there are violations of the minimum social distancing established by the government bodies (Ahmed, et al., 2021). This paper aims to giving a viable contribution to real-time techniques for evaluating interpersonal distancing and detecting whether users wear face masks correctly on public transportation systems both at stations and on-board vehicles. After having designed, calibrated and validated a deep learning-based
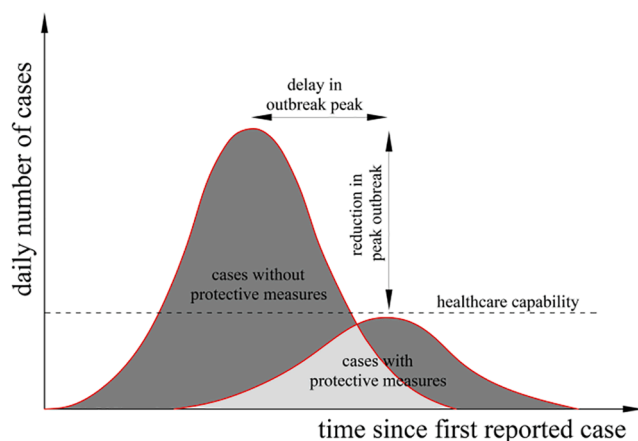

Fig. 1. Qualitative evolution of the number of daily cases of infected people without and with non-pharmaceutical protective measures (i.e. social distancing and face mask wearing).

model, several experiments have been conducted in the public transportation system of an Italian city using low-cost devices. The paper is organized as follows. Sect. 2 explains the related work on object detection and recognition systems based on the deep learning approach and YOLO v3 algorithm. *Sec*. 3 briefly explains the main characteristics of the datasets used for users' detection, the neural network training process and the outputs. Sect. 4 explains the procedure applied for tracking users in the consecutive frames of the surveillance video by means of the Inverse Perspective Mapping (IPM) and the Kalman filter. Sect. 5 includes the description of the case study, the experimental activities and the main results. Finally, conclusions are given in Sect. 6.

Deep learning-based technique

Deep learning (DL) constitutes a subset of artificial intelligence (AI) and is now used in many real-time technical applications for human life facilitation, including the detection of people in an image or video sequences of interest. In general, transportation system user and face mask detection can be considered as object detection and recognition problems (Ottakath, et al., 2022). DL-based models can be efficiently applied for face mask detection and social distance measurement in surveillance systems (Farman, et al. 2022). As demonstrated in the following sections, a video stream can be processed and the DL algorithm is able to detect face masks and the people who violate the social distancing regulations. DL provides end-to-end features extraction process, but in general it necessitates a vast quantity of training data and high computational power. Several research have proved that wearing face masks in public places and maintaining the minimum social distance significantly reduce the spreading rate of COVID-19 (Tang et al., 2009). During the pandemic phase a lot of techniques based on AI procedures were proposed for improving safety in public transportation systems against COVID-19 spread. In general, detection methods can be classified in hand-crafted feature-based methods (i.e. conventional methods) and neural network-based methods, as shown in Fig. 2 (Wang, et al. 2021). Nowadays, Neural Network-based methods (NNBMs) are the most applied techniques for masked face detection. NNBMs can be classified into three main categories: single-stage methods, two-stage methods and multi-stage methods. The latter mainly focus on selective region proposals strategy via a very complex architecture; instead, single stage detectors focus on all the spatial region proposals for the detection of objects via a relatively simpler architecture in one shot (Tausifa, et al., 2022). Nowadays, single stage object detection procedures are significantly better in comparison with most of the two stage object detector methods (Tausifa, et al., 2022). In computer vision (CV) applications, the most significant DL architectures are artificial neural networks (ANNs), convolutional neural networks (CNNs) and adversary generative networks or generative antagonist networks (GANs). Detection systems such as YOLO (You Only Look Once), SSD (single-shot detector) and Faster R-CNN (convolution neural network) are able to identify and classify the objects of interest present in an image, even in scenes characterized by high complex information.

The algorithms of the YOLO families detect the objects of interest by dividing the image into grid cells. Several studies have shown that the algorithms of the YOLO family are very effective in detecting people in the context of an image or video sequence and are often preferred to other types of algorithms (Faster RCNN, SSD, etc.). The second generation of YOLO, called YOLOv2 increases the detection accuracy by utilizing batch normalization to convolution layer, anchor box, multi-scale training and fine-grained features (Lu, et al. 2022). Despite these advantages, the detection accuracy of YOLO v2 is very low for small objects.

Consequently, YOLOv2 is not appropriate for face mask detection. The third generation of YOLO, called YOLOv3, includes 53 convolutional levels and 23 residual levels. YOLOv3 is able to detect in real time the "objects" of interest in a given scene, even if the objects are very small (Guerrieri and Parla, 2021). YOLOv3 performs multi–scale

Fig. 2. State-of-the-art detection methods for people and face mask detection (adapted from: Wang et al. 2021).

classification employing various logistic classifiers. In YOLOv3 the small feature maps give semantic information, instead large ones give finer-grained information. In short, the YOLOv3 algorithm has the qualities of multiscale prediction and multi-label classification. Due to these peculiar characteristics, in the research presented here, the YOLOv3 algorithm, whose schematic architecture is represented in Fig. 3, was used as a fast real-time algorithm for face masks and social distance detection and measure. YOLOv3 utilizes Darknet–53 as the core inter-connection interface, which acts as the attribute extractor for classification. Darknet–53 is characterized by 53 fully convolution layers with residual conjunctions (cf. Fig. 4).

In YOLOv3 the input image is subdivided into S × S grids. Each grid cell predicts three bounding boxes. The method of predicting the bounding box (cf. Fig. 5) is given by the relationships (1):

$$
\begin{cases}
b_x = \sigma(t_x) + c_x \\
b_y = \sigma(t_y) + c_y \\
b_w = p_w e^{t_w} \\
b_h = p_h e^{t_h}
\end{cases}
\tag{1}
$$

In which t_x and t_y represent the relative position coordinates of the centre of the bounding box, t_w and t_h are the width and height of the bounding box. c_x and c_y represent the net, and p_w and p_h are the width and height of the predicted front bounding box. Finally, b_x, b_y, b_w and b_h are the true coordinates of the centre of the bounding box and the true

Fig. 4. Darknet-53 framework in YOLOv3.

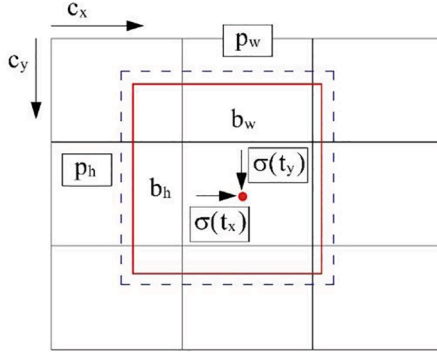**Fig. 3.** YOLOv3 Network structure.

Fig. 5. Bounding Box with dimensions priors and location prediction (adapted from Dewi et al., 2020).

width and height of the bounding box obtained after prediction.

The model predicts a confidence score (t_0) representing the probability that the grid will detect an object for each bounding box by logistic regression (Wan et al., 2021):

$$\sigma(t_0) = P_{r(object)} \ x \ IOU(b, \ object) \tag{2}$$

The probability is 1 when the prediction bounding box coincides with the ground truth location to the cut-off threshold or else the probability value is 0 (Wan et al., 2021).

In order to evaluate the performance of the model in detecting the users of the public transportation system and face masks, metrics such as Precision, Recall and Accuracy were used in this research. The metrics were determined with the following expressions:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

In which TP, TN, FN and FP variables identify the true positives, true negatives, false negatives and false positives respectively. The TP are referred for objects that were labelled true and predicted as true. The TN are for objects that were labelled true but predicted as false. FP are referred for objects that are labelled as false but predicted as false; FN are the images that are labelled false but predicted as true (Ottakath, et al. 2022).

One of the most important criteria for evaluating the performance of the detection model is the loss function. In Yolov3 the loss function is divided into three main components: classification loss, localization loss (errors between the predicted boundary box and the ground truth) and confidence loss (the objectness of the box). The loss function can be expressed as follows (Hui, 2018):

Table 1
Available open-source datasets.

Dataset Name	Available Link Access Date
MAFA	https://drive.google.com/drive/folders/1nbtM1n0–iZ3VVbNGhocxbnBGhMauOG
MFDD	https://github.com/X-zhangyang/ Real-World-Masked-Face-Dataset
MFNID	https://github.com/cabani/MaskedFace-Net
Moxa3K	https://shitty-bots-Inc.github.io/MOXA/index.html
PWMFD	https://github.com/eth ancvaa/Properly-Wearing-Masked-Detect-Dataset April
UFMD	https://github.com/iremeyiokur/COVID-19-Preventions-Control-System
FMLD	https://github.com/borutb-fri/FMLD
Dey Dataset	https://github.com/chandrikadeb7/Face-Mask-Detection
Singh Dataset	https://drive.google.com/drive/folders/1pAxEBmfYLoVtZQlBT3doxmesAO7n3ES1?usp=sharing
WMD	https://github.com/BingshuCV/WMD
AIZOO -Tech	https://github.com/AIZOOTech/FaceMaskDetection
Kaggle	https://www.kaggle.com/andrewmvd/face-mask-detection
SMFD	https://github.com/prajnasb/observations
COVID-19 Medical Face Mask Detection Dataset	https://www.kaggle.com/datasets/mloey1/medical-face-mask-detection-dataset

Training of the neural network

There are two key problems in face mask wearing detection: the former is to locate the position of a user's face in the image under analysis and the latter is to identify whether the face mask is worn correctly. Therefore, data collection is of paramount importance for neural network training. The most important open-source public datasets are summarised in Table 1. Since the proposed model must be able to detect users of public transportation systems and facial masks, in this research two different datasets were used for training the neural network: the "Caltech Pedestrian Dataset" (Fig. 6) and the "COVID-19 Medical Face Mask Detection Dataset" (Fig. 7). The Caltech Pedestrian Dataset is based on about ten hours of video recordings (with resolution 640x480, 30 Hz) from which about 250,000 frames were extracted, and 350,000 bounding boxes and 2300 pedestrians were identified. For the purposes of the study described here, 1000 images of pedestrians were extrapolated from the Caltech Pedestrian Dataset and 800 images were taken from the "COVID-19 Medical Face Mask Detection Dataset". There are two categories of images in the training data collections: with mask and without mask. To properly train the neural networks the total image dataset was split into two different sets: 70 % of the image sample was used for neural network training and 30 % for testing. It should be noted that the labelling of the ground truth bounding box (ground truth) was done "manually". The result of the analyses was that the neural network has "learned" to recognise users of the public transportation system with

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=o}^{B} I_{ij}^{obj} [(x - \hat{x}_i)^2 + (y - \hat{y}_i)^2] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=o}^{B} I_{ij}^{obj} [(\sqrt{\varpi_i} - \sqrt{\hat{\varpi}_i})^2 + \sqrt{h_i} - \sqrt{\hat{h}_i})^2] +$$

$$\sum_{i=0}^{S^2} \sum_{j=o}^{B} I_{ij}^{obj} (C - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=o}^{B} I_{ij}^{noobj} [(C - \hat{C}_i)^2 + \sum_{i=0}^{S^2} I_{ij}^{obj} (p_i(c) - \hat{p}_i(c))^2 \tag{6}$$

Fig. 6. Sample images from the "Caltech Pedestrian Dataset".

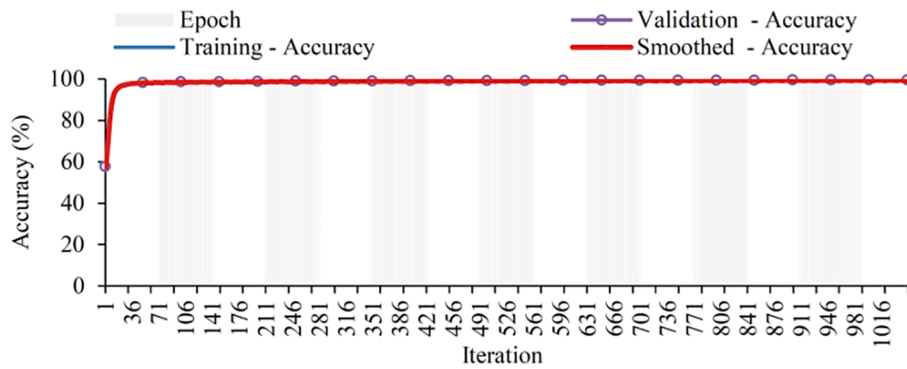**Fig. 7.** Sample images from the "COVID-19 Medical Face Mask Detection Dataset".

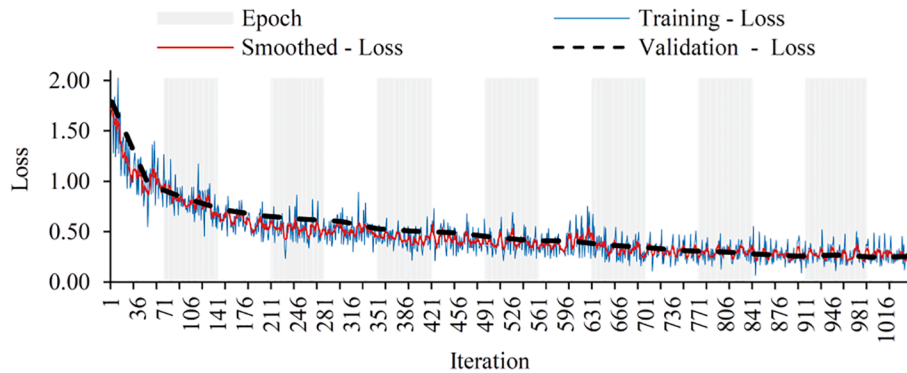**Fig. 8.** Accuracy evolution related to the number of iterations.

Fig. 9. Loss evolution related to the number of iterations.

or without protective face masks.

The neural network training process was performed using a workstation with Intel(R) Core(TM) i7-4510 CPU @ 2.00 Hz 2.60 GHz – Memory RAM 6 GB, Windows 10 Home. In all, the process requested a Time Elapsed of 00:50:30 h and 60 Epoch (1 Epoch = 25 iterations); the following values were assumed: learning rate = 0.001; L2 regularization factor = 0.0005, penalty threshold = 0.5.

Starting from a surveillance video, for each processed image, the algorithm evaluates:

- the coordinates of the position of the Bounding Box related to each user of the public transportation system;
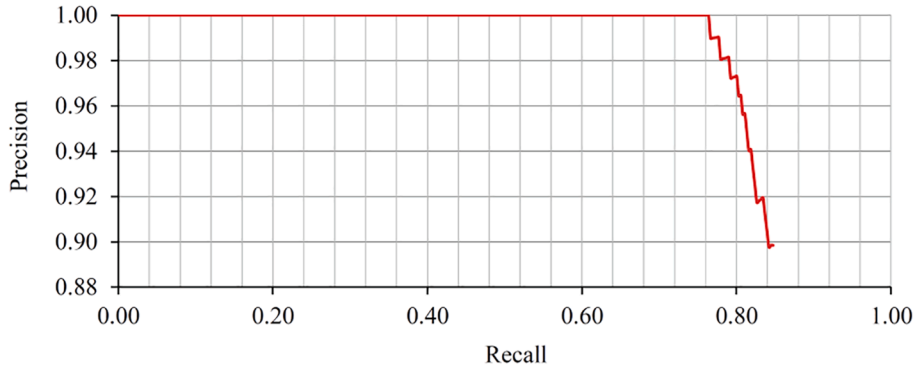
Fig. 10. Precision-Recall curves.

- the coordinates of the position of the Bounding Box related to each face mask;
- the statistical confidence level associated with each Bounding Box;
- the label that identifying the class name of the identified object (user of the transportation system and face mask).

Figs. 8–10 illustrate the Loss, Accuracy and Precision values obtained in the testing process; it is clear that when accuracy increases, loss decreases.

Tracking of users and social distancing evaluation

Starting from the analysis of a video recording, each user of a given public transportation system is detected with the procedure explained in the previous sections. Reliable user tracking in the different frames of a video recording can be achieved by using the Kalman filter. Consequently, for each instant of time the proposed technique identifies clusters of people who maintain a social distance (d) lower than the minimum social distance rule (d_{min}) and evaluates the exposure time. In this research, the estimated social distance is that one between the people centroids projected on the ground plane (i.e. road pavement surface for stop areas or vehicle floor) by the inverse perspective mapping (IPM) procedure (Dorj & Lee, 2016). With the use of IPM the position of the bounding box of each detected user in the perspective view is converted into a top-down view (Fig. 11). Denoting with:

- $\{F_w\} = \{X_w, Y_w, Z_w\}$ the world frame centered at the camera optical center;
- $\{F_c\} = \{X_c, Y_c, Z_c\}$ the camera frame;
- $\{F_i\} = \{u, v\}$ an image frame (Fig. 19);

- h the height of the camera frame (Fig. 9) with respect to the ground plane,

under the hypothesis that the optical axis has no roll (i.e. the camera frame X_c axis is in the world frame $X_w Y_w$ plane), the projection point on the road plane of each point $^iP = \{u, v, 1, 1\}$ on the image plane can be estimated by Eq. (7) (Aly, 2008):

$$
{}^g_i T = h
\begin{bmatrix}
-\dfrac{1}{f_u}c_2 & -\dfrac{1}{f_v}s_1 s_2 & \dfrac{1}{f_u}c_u c_2 - \dfrac{1}{f_v}c_v s_1 s_2 - c_1 s_2 & 0 \\[2ex]
\dfrac{1}{f_u}s_2 & \dfrac{1}{f_v}s_1 c_1 & -\dfrac{1}{f_u}c_u s_2 - \dfrac{1}{f_v}c_v s_1 c_2 - c_1 c_2 & 0 \\[2ex]
0 & \dfrac{1}{f_v}c_1 & -\dfrac{1}{f_v}c_v c_1 + s_1 & 0 \\[2ex]
0 & -\dfrac{1}{hf_v}c_1 & \dfrac{1}{hf_v}c_v c_1 - \dfrac{1}{h}s_1 & 0
\end{bmatrix}
\tag{7}
$$

where:

- $\{f_u, f_v\}$ are the horizontal and vertical focal lengths respectively;
- $\{c_u, c_v\}$ are the coordinates of the optical centre;
- $c_1 = \cos\alpha$, $c_2 = \cos\beta$, $s_1 = \sin\alpha$, and $s_2 = \sin\beta$.

Similarly, given a point on the road plane $^gP = \{x_g, y_g, -h, 1\}$, its subpixel coordinates can be calculated on the image frame by the equation $iP = {}^i_g T \, {}^gP$, i.e. by the inverse of the transform (Aly, 2008):

$$
{}^g_i T =
\begin{bmatrix}
f_u c_2 + c_u c_1 s_2 & c_u c_1 c_2 - s_2 f_u & -c_u s_1 & 0 \\
s_2 (c_v c_1 - f_v s_1) & c_2 (c_v c_1 - f_v s_1) & -f_v c_1 - c_v s_1 & 0 \\
c_1 s_2 & c_1 c_2 & -s_1 & 0 \\
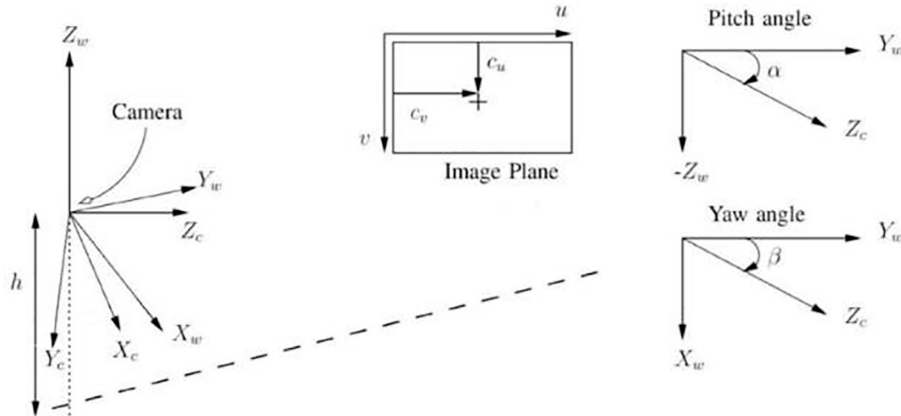c_1 s_2 & c_1 c_2 & -s_1 & 0
\end{bmatrix}
\tag{8}
$$

Fig. 11. IPM coordinates. Left: coordinate axes (world). Right: definition of pitch and yaw angles.
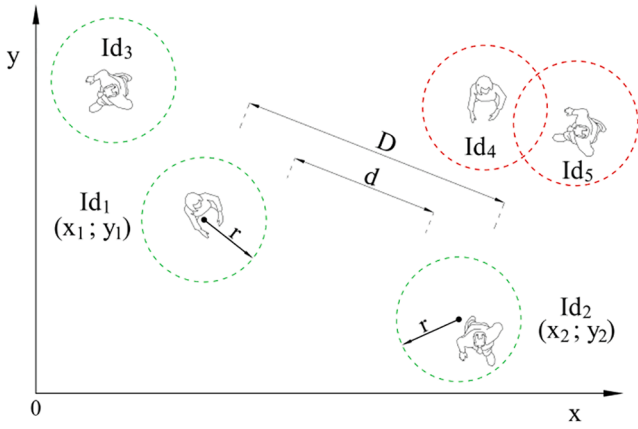
Fig. 12. Evaluation of the social distance violation in function of the Euclidean distance (the red circles indicate the minimum social distance violation). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A Bird's-eye view of users' centroids can be obtained from Eq. (7). Consequently, we can project the centroids of all users of the public transportation system from the input image onto the ground plane and then calculate the social distances. In fact, for each user identified in the scene of each frame of the video under analysis, the model creates a circumference of radius r equal to half the minimum social distance to be respected (d_{min}) (Fig. 12). Denoting with (x1; y1) and with (x2; y2) the coordinates of the centroids of two detected people (Id1 and Id2 respectively, cf. Fig. 12) the social distance violation occurs if $d < d_{min}$.

The Euclidean distance (d) can be evaluated by Eqs. (9), (10), (11):

$$d = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} - 2r = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} - 2\frac{d_{min}}{2}$$
$$= \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} - d_{min} \tag{9}$$

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d_{min} \geqslant 0 \rightarrow \text{no social distance violation} \tag{10}$$

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} - d_{min} < 0 \rightarrow \text{social distance violation} \tag{11}$$

In the final output the social distancing violations are automatically identified by red circles in a "bird's eye view" of the scene and by red bounding boxes in the original images.

Since one or more users can be detected in successive frames of the analysed video, it is crucial to apply a tracking algorithm. The methodology used in this work is summarised in Fig. 13. In particular, when a user of the public transportation system comes into the scene, it is considered as a new tracking object, then the algorithm assigns the Id number and initializes the bounding box for that user. To decrease the noise introduced by inaccurate or imprecise detections the linear Kalman filter is employed. As it is well known, the Kalman filter (Kalman, 1960) is a recursive predictive filter that assesses the status of a dynamic system. The Kalman filter equation is (Welch and Bishop, 2006):

$$x_{n+1} = A_n x_n + B_n u_n \tag{12}$$

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

$$P_{n+1} = A_n P_n A_n^T + Q_n \tag{13}$$

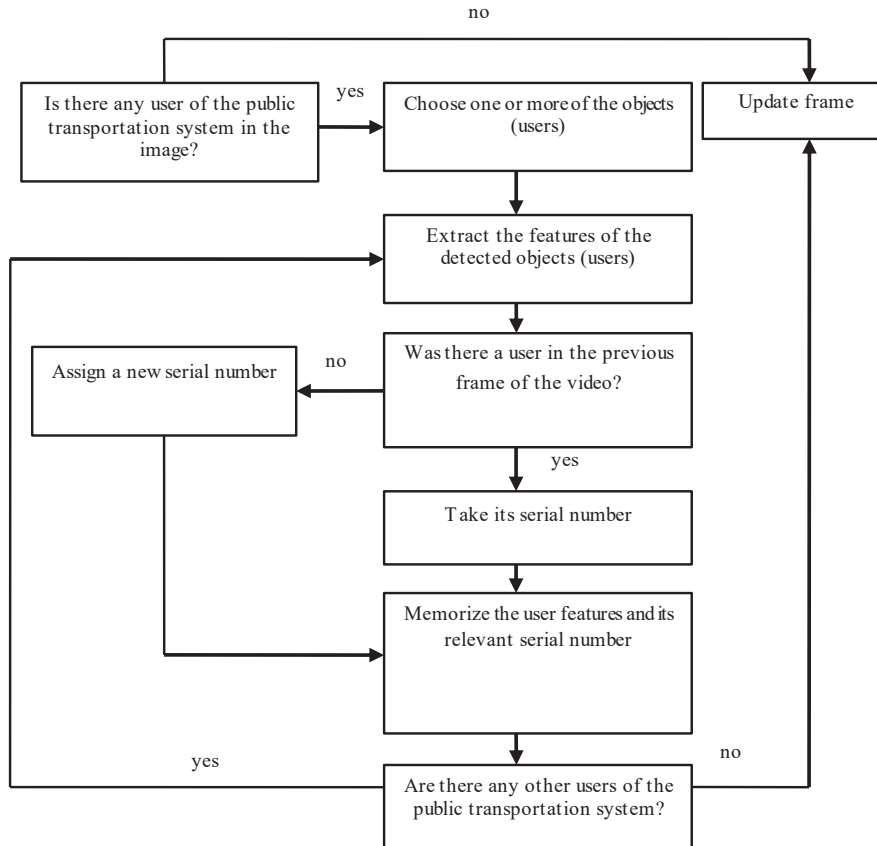where x_n is the state value at step n, A_n is the state transition matrix, u_n is

Fig. 13. Tracking algorithm operations for detecting a user of the public transportation system in the video sequence.

the measurement and the input at step n. Q_n is the white noise covariance. Kalman gain value is given by the following relationship (Niu, 2018):

$$K_n = P_n C^T (C P_n C^T + R_n)^{-1} \tag{14}$$

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

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

$$P_n = (I - K_n H) P_n \tag{15}$$

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

An example of user trajectory calculated with the Kalman filter is given in Fig. 16a.

Experimental activities and results

The complete description of the experiment executed in this research is given in this section. The experimental activities were undertaken in January and February 2022 in the Public Transportation System (PTS) of the Municipality of Piazza Armerina (Italy), operated by the company Savit Scichilone Ltd. In particular, some video cameras were installed at stops and inside buses (Fig. 9). The recorded video streams were split up into frames of sequences for the accurate estimation of the distance values between each couple of the detected users.

The most important challenge of detecting social distancing from videos is the accuracy of the measurement of the actual distance between public transportation users (Zuo, et al., 2021). To solve this problem, the holography technique can be applied to morph the video frames from a perspective view into a top–bottom view. Considering prefixed objects with known dimensions as references, distances can be computed in the transformed frames (Szeliski, 2010). Therefore, the first phase of the experimental activities concerned the camera calibration so as to obtain the extrinsic parameters of the cameras themselves (Fig. 14). To this end, Zhang's algorithm was used and applied to a set of twenty-two different images of a chessboard of known size. The chessboard is employed to scale the pixels in the image acquired by the camera to compute the focal length, the principal point etc (Gad, et al. 2020). The camera calibration parameters, including the height from the ground level in the scene, the yaw and the pitch, are used in the inverse perspective mapping (IPM) process to create a bird's eye view of the images of interest. In conclusion, by using the procedures set out in the previous sections, each user of the PTS is detected within a certain video recording. The original input frames are subject to a Gaussian smoothing filter to reduce noises in the images. Subsequently, an order number is assigned to each user and the position of his centroid in the coordinate system centred in the photographic objective, and the mutual distances are calculated between the users identified in the scene. The system then checks whether the minimum social distancing between single pairs of users, or between different pairs of users belonging to the same cluster, is
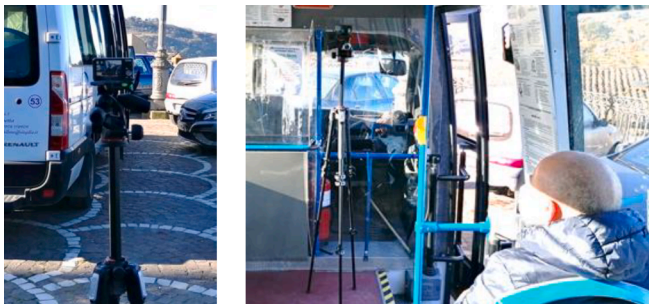

Fig. 14. Location of the cameras in stop areas and inside buses of the PTS of Piazza Armerina municipality.
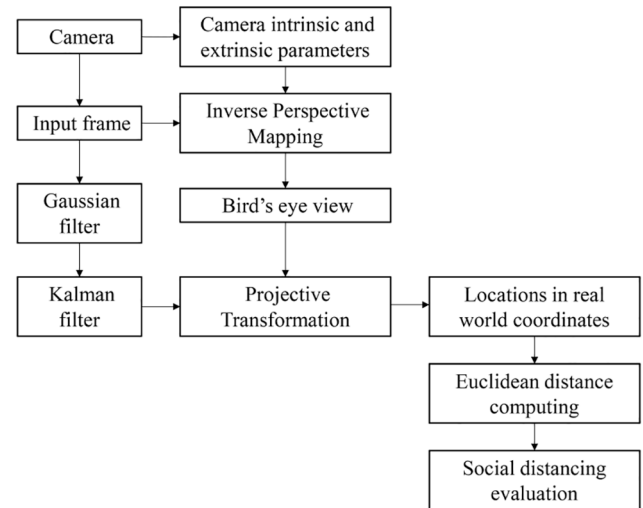

Fig. 15. Flow chart of the overall proposed technique for social distancing evaluation.

respected or not. The overall procedure is summarised in Fig. 15.

This procedure is applicable for analysing safety condition external to the vehicles of the PTS (i.e. at stop areas as shown in Figs. 16 and 17 or for users entering and exiting vehicles) and those inside vehicles (Fig. 18). As a result of the analyses performed by the proposed algorithm, the violations of the minimum social distancing are highlighted with red bounding boxes and can also be correlated to the time of exposure to the risk of infection, that is the interval of time during which the minimum social distancing is violated. Finally, the proposed technique is able to identify whether users are wearing respiratory protection FFP2 face masks correctly or not (Fig. 19). As can be observed in Figs. 18 and 19, users maintaining the minimum permissible social distance are identified using green bounding boxes, whereas the users violating the threshold limit for social distancing are highlighted by red bounding boxes or circles. Fig. 17c shows an example of coordinates for user centroids obtained by a bird's-eye view (cf. transformation matrix given in Eq. (7) and Eq. (8)) for a specific instant of time. After determining the coordinates (in pixels) of each user's centroid, the pixels in a bird's-eye view are converted into metres in order to compute the social distancing between a couple of users.

As part of the experiments only few cases of false positives and negatives were generated, but in general the results have made it possible to highlight a considerable reliability of the proposed method in detecting real time interactions between users of the PTS in terms of over-time variations in their mutual distancing, as well as in recognising cases of violation of the imposed rules. The number of false negatives and false positives decreases when the longitudinal, lateral and vertical acceleration of the camera decreases during the vehicle motion. An increase in camera oscillation can be observed near horizontal curves (Vaiana et al., 2018), road intersections (Gallelli, et al., 2019; Mauro and Guerrieri, 2016; Guerrieri, et al. 2018) or in case of road pavement damages. Therefore, in order to increase the precision of the proposed technique it is essential to usevibration reduction lenses and vibration sensor technologies. These technologies help reduce the amount of blur on a video caused by movement of the camera due to vehicle oscillations with consequent benefits in terms of accuracy in object detection. Experimental outcomes have demonstrated that the proposed technical approach is characterised by notable accuracy and detection speed and can effectively detect the users of the public transportation systems who maintain or breach the permitted minimum social distance as well as calculate the number of violations. In addition, the Loss, the Accuracy and the Precision values achieved in the training process of the neural network and the experimental results show that the algorithm proposed
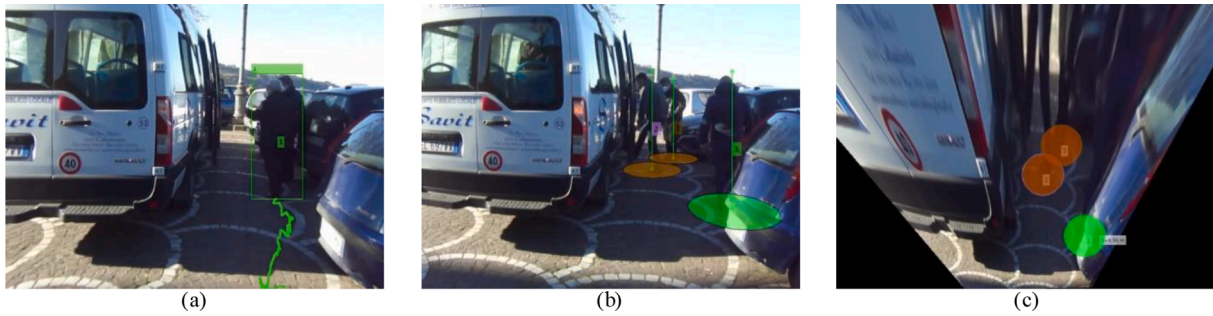
Fig. 16. a) Example of a user's trajectory obtained using Kalman filter. b) Example of minimum social distance violation (red circles); c) Example of a Bird's Eye view. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
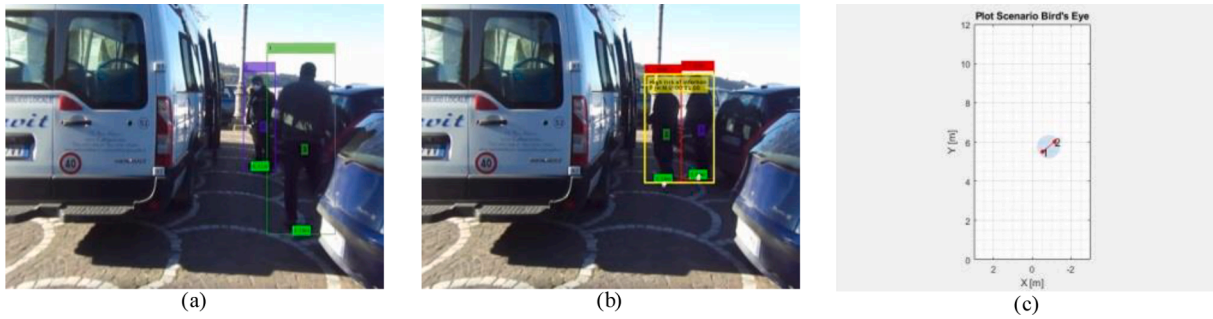

Fig. 17. a-b) Detection of users in a stop area of the PTS; c) social distance evaluation.

Fig. 18. Detection of users inside a vehicle of the PTS and social distance measurement.

Fig. 19. Detection of users' face masks (FFP2 type) inside a vehicle of the PTS and social distance measurement.

has high accuracy in face mask detection.

Conclusions

This article presents a model based on computer vision, deep learning and YOLOv3 as a contribution to preserve a safe environment by scanning the areas inside vehicles or around the stops of a public transportation system in order not to let COVID-19 spread out. The technique proposed aims not to prevent the social distance and the use of face mask violations but rather to decrease their excessive occurrence in public transportation systems. Such a technique, jointly with a user's face recognition algorithm, can speed up the process of contact tracing that is a key strategy for breaking chains of transmission of SARS-CoV-2 and decreasing COVID-19-associated mortality. In addition, the procedure can be associated with visual and acoustic real-time signals that immediately alert users whenever they are about to breach the

mandatory minimum distance from other people. In this research the neural network was trained by using a sample of images of the "Caltech Pedestrian Dataset" and the "COVID-19 Medical Face Mask Detection Dataset" were used. Although the proposed method still requires a vast experimental campaign for properly validating the algorithms, the first results achieved in the case study of the public transportation system (PTS) of the municipality of Piazza Armerina (Italy) showed that the technique is able to accurately measure the reciprocal distance between users and to detect the presence of face masks (type FFP2). In fact, face detectors perform very well even if their accuracy may be further improved. Such a technique can also track both transient and steady-state people in order to measure social distance at each instant of time both in stop areas and on-board vehicles. The outputs of the model, in terms of bounding boxes, help in identifying couples or groups of people that satisfy the required minimum social distance and the number of violations, also correlated to the time of exposure to the risk of infection. It is worth underlining that the methodological approach suggested in the study, although limited only to one PTS, is general and can be applied to other urban public transportation systems as well as railways, metro, light rail systems etc.

CRediT authorship contribution statement

Marco Guerrieri: Conceptualization, Methodology, Writing– original draft, Writing – review & editing. **Giuseppe Parla:** Investigation, Data curation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

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.

Aly, M., 2008. Real time Detection of Lane Markers in Urban Streets. IEEE Intelligent Vehic. Symposium, Proc. 7–12.

Calderón Peralvo, F., Cazorla Vanegas, P., Avila-Ordóñez, E., 2022. A systematic review of COVID-19 transport policies and mitigation strategies around the globe. Transp. Res. Interdiscip. Perspect. 15, 100653.

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, 2013, 59(1), pp. 3–19.

De Vos, J., 2020. The effect of COVID-19 and subsequent social distancing on travel behavior. Transp. Res. Interdiscip. Perspect. 5, 100121.

Dewi, C., Chen, R.-C., Yu, H., 2020. Weight analysis for various prohibitory sign detection and recognition using deep learning. Multimedia Tools Appl. 79 (43–44), 32897–32915.

Dorj, B., Lee, D.J., 2016. A Precise Lane Detection Algorithm Based on Top View Image Transformation and Least-Square Approaches. J. Sens. art. no. 4058093.

Eksin, C., Paarporn, K., Weitz, J.S., 2019. Systematic biases in disease forecasting–the role of behavior change. Epidemics 27, 96–105.

Farman, H., Khan, T., Khan, Z., Habib, S., Islam, M., Ammar, A., 2022. Real-Time Face Mask Detection to Ensure COVID-19 Precautionary Measures in the Developing Countries. Appl. Sci. (Switzerland) 12 (8).

Gad, A., Elbary, G., Alkhedher, M., Ghazal, M. Vision-based Approach for Automated Social Distance Violators Detection (2020). 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies, 3ICT 2020, art. no. 9311969.

Gallelli, V., Guido, G., Vitale, A., Vaina, R., 2019. Effects of calibration process on the simulation of rear-end conflicts at roundabouts. J. Traffic Transp. Eng. (English Edition). 6 (2), 175–184.

Grant, M.J., Booth, A., 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf. Libr. J. 26, 91–108. https://doi.org/10.1111/J.1471-1842.2009.00848.X.

Guerrieri, M., 2019. Catenary-Free Tramway Systems: Functional and Cost-Benefit Analysis for a Metropolitan Area. Urban Rail Transit 5 (4), 289–309.

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, 2018, 144(6), 04018020.

Guerrieri, M., Parla, G., 2021. Deep learning and yolov3 systems for automatic traffic data measurement by moving car observer technique. Infrastructures 6 (9).

Hui, J. 2018. Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3. Available: https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088.

Kalman, R.E., 1960. A new approach to linear filtering and predictions problems. J. Basic Eng. 82 (D), 35–45.

Kermack, W. O., McKendrick, A. G. 1927. Contributions to the mathematical theory of epidemics–i.

Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J., 2019. The incubation period of coronavirus disease (CoVID-19) from publicly reported confirmed cases: Estimation and application. Ann. Internal Med. 172 (9), 577–582.

Lu, H., Chen, T., Shi, L., 2022. Research on Small Target Detection Method of Traffic Signs Improved Based on YOLOv3. 2022 2nd International Conference on Consumer Electronics and Computer Engineering, ICCECE.

Mauro, R., Guerrieri, M., 2016. Comparative life-cycle assessment of conventional (double lane) and non-conventional (turbo and flower) roundabout intersections. Transp. Res. Part D: Transp. Environ. 48, 96–111.

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 accessed on 10 August 2021).

Ottakath, N., et al. 2022. ViDMASK dataset for face mask detection with social distance measurement. Displays, 73, art. no. 102235.

Park, J., Kim, G., 2021. Risk of COVID-19 Infection in Public Transportation: The Development of a Model. Int. J. Environ. Res. Public Health. 18, 12790. doi: 10.3390/ijerph182312790.

Prather, K.A., Wang, C.C., Schooley, R.T., 2020. Reducing transmission of SARS-CoV-2. Science 368 (6498), 1422–1424. https://doi.org/10.1126/science.abc6197.

Szeliski, R., 2010. Computer Vision: Algorithms and Applications. Springer Science & Business Media.

Tang, J.W., Liebner, T.J., Craven, B.A., Settles, G.S., 2009. A schlieren optical study of the human cough with and without wearing masks for aerosol infection control. J. R. Soc. Interface 6, S727–S736.

Tausifa, S., Anirudh, G., Tembhurne, J., 2022. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimedia Tools Appl. https://doi.org/10.1007/s11042-022-13644-y.

Vaiana, R., Iuele, T., Gallelli, V., Rogano, D., 2018. Demanded versus assumed friction along horizontal curves: An on-the-road experimental investigation. J. Transp. Saf. Security 10 (4), 318–344.

Wan, J., et al., 2021. An Efficient Small Traffic Sign Detection Method Based on YOLOv3. J. Signal Process. Syst. 93 (8), 899–911.

Wang, B., Zheng, J., Philip Chen, C.L. 2021. A Survey on Masked Face Detection Methods and Datasets for Fighting Against COVID-19. Journal of IEEE transactions on artificial intelligence, 0(0), December 2021.

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 (accessed on 1 August 2021).

Zuo, F., Gao, J., Kurkcu, A., Yang, H., Ozbay, K., Ma, Q. 2021. Reference-free video-to-real distance approximation-based urban social distancing analytics amid COVID-19 pandemic. Journal of Transport and Health, 21, art. no. 101032.