Social Distancing Violation Detection in Video Using ChessBoard and Bird's-eye Perspective

An Cong Tran, Trong Huu Ngo, and Hai Thanh Nguyen[✉]

Can Tho University, Can Tho, Vietnam
`tcan@cit.ctu.edu.vn`, `nthai.cit@ctu.edu.vn`

Abstract. Thousands of infections, hundreds of deaths every day - these are numbers that speak the current serious status, numbers that each of us is no longer unfamiliar with in the current context, the context of the raging epidemic - Coronavirus disease epidemic. Therefore, we need solutions and technologies to fight the epidemic promptly and quickly to prevent or reduce the effect of the epidemic. Numerous studies have warned that if we contact an infected person within a distance of fewer than two meters, it can be considered a high risk of infecting Coronavirus. To detect a contact distance shorter than two meters and provides warnings to violations in monitoring systems based on a camera, we present an approach to solving two problems, including detecting objects - here are humans and calculating the distance between objects using Chessboard and bird's eye perspective. We have leveraged the pre-trained InceptionV2 model, a famous convolutional neural network for object detection, to detect people in the video. Also, we propose to use a perspective transformation algorithm for the distance calculation converting pixels from the camera perspective to a bird's eye view. Then, we choose the minimum distance from the distance in the determined field to the distance in pixels and calculate the distance violation based on the bird's eye view, with camera calibration and minimum distance selection process based on field distance. The proposed method is tested in some scenarios to provide warnings of social distancing violations. The work is expected to generate a safe area providing warnings to protect employees in administrative environments with a high risk of contacting numerous people.

Keywords: Minimum distance · Bird's eye view · Human recognition · Epidemic · Distancing violation

1 Introduction

Artificial intelligence is present in all areas of human life, from economics, education, and medicine to housework, entertainment, and even the military. But the breakthroughs mostly come from Deep Learning - a small array gradually expanding to each type of work, from simple to complex, like in the field of positioning with tasks that need high accuracy, including drones and uncrewed

vehicles. In the context that the world is facing COVID-19, with thousands of people infected every day, preventing the spread is the issue that needs the most attention, so distance violation detection approaches through using the camera have become useful. As a result, we can detect and timely prevent the spread of the disease or create favorable conditions. In addition, the technologies tracing the origin of the spread have been proposed and developed so that disease prevention and control take place effectively, promptly, and quickly.

Public places and crowded places such as schools, administrative offices, and hospitals are clear places to spread Covid pathogens. In such places, keeping a distance is one of the important rules in the 5K rule [1][1] of Ministry of Vietnamese Public Health released, including wearing a mask, disinfection, not gathering, declaring health and keeping the minimum distance. Keeping a distance is extremely necessary during the pandemic to prevent the spread of disease. Health experts say that keeping a distance is one of the most important and recommended solutions to prevent the pandemic. However, not everyone understands the role and meaning of distance and considers it a serious rule. Coronavirus disease, as we know, is transmitted mainly through close contact within about two meters. Therefore, there is a very high infection risk, as the Vietnamese Ministry of Health recommended in 2020. Moreover, numerous evidences from research was restated that the 2-meter social distancing rule to reduce COVID-19 transmission [2,3]. Although we know that the advice on 2-meter distancing is a risk assessment of some research, such distance should be longer to ensure safety. Also, the authors in an article [4] in the Lancet journal stated that physical distancing of at least one meter lowers the risk of COVID-19 transmission, but that two meters can be more effective. Exposure occurs when droplets from the nose or mouth of an infected person cough, sneeze, or talk shoot out and into the air, and these elements come into contact with an uninfected person that can cause illness. Many studies have shown that asymptomatic infected people also contribute to the spread of COVID-19 because they can spread the virus before they reveal symptoms.

This study proposes an approach to detect minimum distance violations by the camera to support the implementation of regulations on keeping a distance of two meters when in contact with others. The rest of the work is organized as follows. First, we discuss some related applications and studies in Sect. 4. After that, in Sect. 3, we will elaborate on our workflow and algorithms. Subsequently, our experiments with test cases will be described and explained (Sect. 4). Finally, in conclusion, we will summarize our studies' key features and development plan (Sect. 5).

2   Related Work

With the serious effects of Corona disease, numerous scientists have focused on some solutions to prevent the epidemic in a vast of applications and studies.

---

[1] https://covid19.gov.vn/bo-y-te-khuyen-cao-5k-chung-song-an-toan-voi-dich-benh-1717130215.htm.

Some studies have attempted to present computer-based methods to monitor social distancing violations. The work in [5] used the YOLOv3 object recognition algorithm to indicate humans in video sequences with a pre-trained algorithm to gather to an extra-trained layer using an overhead human data set. In [6], the authors designed a computer vision-based smart to monitor and automatically detect people who violated safe distancing rules. In another study in [7], the authors evaluated the American people's perceptions of social distancing violations during the COVID-19 pandemic. In work in [8], scientists introduced SocialNet, a method aimed to indicate violations of social distancing in a public crowd scene. The method included two main parts, including a detector backbone and an Autoencoder. The research in [9] deployed the YOLO algorithm to detect social distancing violations in real time. In addition, the authors in [10] presented a drone using a surveillance method implemented in Deep Learning algorithms to indicate whether two people were violating social distancing rules. The authors in [11] used a small neural network architecture to indicate social distancing using a bird's eye perspective. In more studies on deep learning-based method, the study in [12] implemented bounding boxes to indicate group violated the social distancing rules using Euclidean distance and deep learning trained on COCO dataset [13] using YOLO. In [14], the authors introduced a method to detect people in a frame and check whether people violated social distancing by calculating the Euclidean distance between the centroids of the detected boxes. The authors in [15] deployed the "Nvidia Jetson Nano" development kit and Raspberry Pi camera to compute and determine social distance violation cases. The scientists in [16] counted the violations using some analysis techniques on video streams.

Bluezone application[2] provided warnings and contacted people infected with Covid-19. Some outstanding features of the Bluezone application include a scan for nearby Bluezone community, warning when contact with Covid-19 infected person, secure operation, transparency, lightweight, and low battery consumption. In addition, it could support people to make medical declarations right on their phone easily at any time, provide a quick electronic medical declaration using a QR code, and easily track contact history with Bluezone users. In addition, it can submit a report of COVID-19 disease summarized information, suspected infected subjects around the area where you live, provide an Electronic health book, and allows Covid-19 vaccination registration. Health Book application[3] was also an interesting application with some features such as allowing people to register for the Covid-19 vaccine on the app. In addition, it can report any unusual symptoms quickly after getting the COVID-19 vaccine, providing a certificate of vaccination against COVID-19. Moreover, it can support declaring health information and family anytime, anywhere. It can easily track health tracking after connecting the COVID-19 vaccine directly with the personal health record system of the Ministry of Health, easily book an appointment with a medical facility or doctor before visiting, or maybe, talk to your doctor

---

[2] https://bluezone.gov.vn/.

[3] https://play.google.com/store/apps/details?id=com.mohviettel.sskdt.

online for advice and take care of your health. The most obvious common difficulty is that it is hard to monitor accurately. However, now that the policy of loosening the distance is applied, more and more people gather in public places, so using a Camera to monitor is essential and popular.

3   Methods

The camera's approach to detecting distance violations is proposed in Fig. 1.
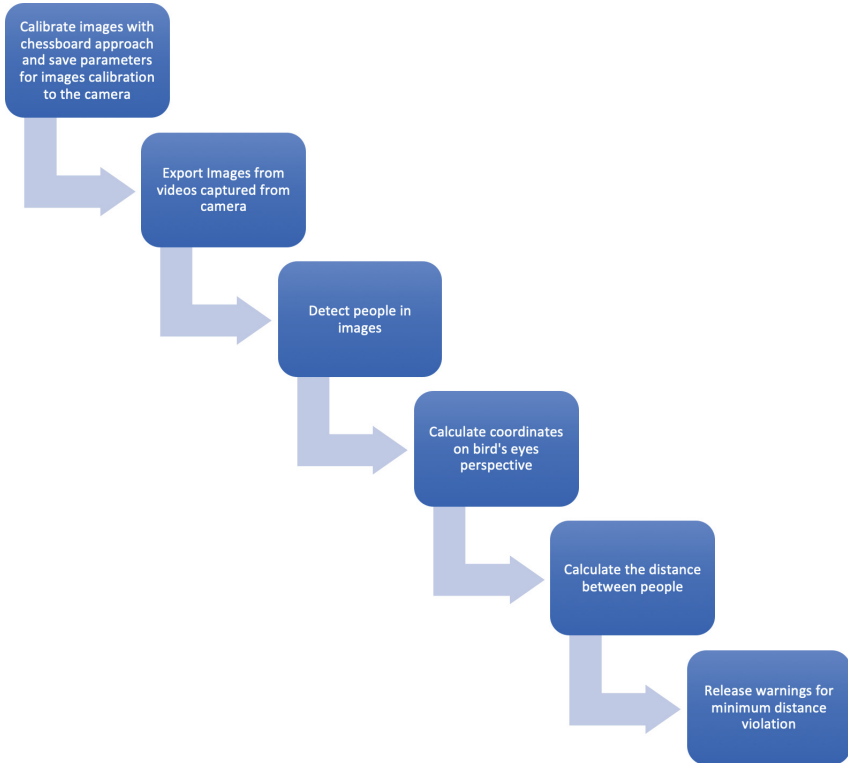
Fig. 1. Steps of the process of building a model to detect distance violations through the camera.

First, we calibrate the Camera with OpenCV to increase the model's accuracy. After this calibration, we can conduct object recognition with the pre-train model to create bounding boxes indicating people's positions. After having these bounding boxes, we can calculate the coordinates of the midpoint of the bottom edge of the bounding box, taking that as a basis to calculate the distance between the objects. Finally, we have deployed the chessboard method to calibrate images from videos captured from the camera's output before converting the image into perspective from camera view to birds'-eye view.

### 3.1   Object Recognition with Pre-trained Inceptionv2 Model

This study has deployed Inceptionv2 [17] to recognize the object recognition process. Inceptionv1 [18] was originally proposed with about 7 million parameters. It was much smaller than the famous prevailing architectures, like VGG [19] and AlexNet [20]. However, it can achieve a lower error rate, which is also why it is a breakthrough architecture. The modules in Inception perform convolutions with different filter sizes on the input, operate max pooling and concatenate the results for the next inception modules. The introduction of a 1×1 convolution operation greatly reduces the parameters. Although the number of layers in Inceptionv1 is 22, the dramatic parameter reduction has made it a very difficult model to overfit. The Inceptionv2 [17] is a major improvement on the Inceptionv1 that increases accuracy and further makes the model less complex. The significant improvements in the Inceptionv2 model include Multiplying the 5×5 convolution into two 3×3 convolution operations to improve the computation speed. Although this may seem counter-intuitive, a 5×5 convolution is 2.78 times more expensive than a 3×3 convolution. So stacking two 3×3 convolutions leads to a performance increase. Furthermore, Inception factorizes filter convolutions of size $n \times n$ into a combination of $1 \times n$ and $n \times 1$ convolutions. For example, a 3×3 convolution is equivalent to first performing a 1×3 convolution and then performing a 3×1 convolution on its output. They found that this method was 33% in time execution reduction than 3×3 single convolution.

3.2   Calibrate the Camera with Chessboard Corner Detection

Cameras these days have become relatively cheap to manufacture. We have deployed camera calibration from captured images to indicate the geometric parameters of the image formation process. As mentioned in [21], the camera calibration process is an important step in computer vision tasks, especially when metric information about the scene is required. We calibrate the camera to enhance the efficiency of computing the distance between two objects with some changes for the internal parameters of the camera (camera calibration change due to movement of the internal lens) even if it falls to the ground and causes production problems) and external parameters are calculated. There are two types of uncorrected in-camera noise. The first is called barrel distortion and causes a quick view from the sides, while the second type is called pin buffer noise (pincushion distortion) and is flattened from the sides.

We call the matrix known as the denoising camera matrix. OpenCV [22] provides calibration support with various methods. The most famous of these is the Chessboard corner detection [23]. With the chessboard corner detection, we first detect corners in the chess board. Then, we draw detected angles using drawChessboardCorners to generate a new image with circles at the corners found in python. Next, we calculate the camera's internal and external parameters from multiple perspectives of a calibrator. Our final step is to store the parameters returned by feeding the feature points in all images and the equivalent pixels in the two-dimensional image into the calibrated camera function.

The total error represents the accuracy of the camera calibration process. Total error calculation is done by projecting 3D checkerboard points into the image plane using the final correction parameters. A Root Mean Square Error of 1.0 means that, on average, each of these projected points is 1.0 pixels from its actual location. The error is not bound in [0, 1]. It can be considered a distance as an illustration in Fig. 2.
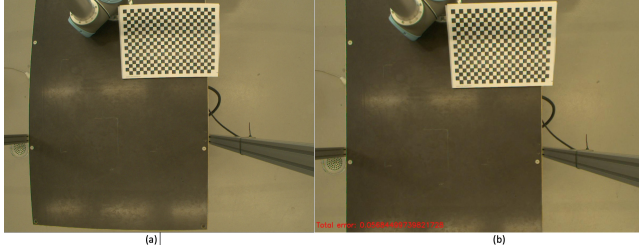
Fig. 2. An original image and its image after calibrating camera.

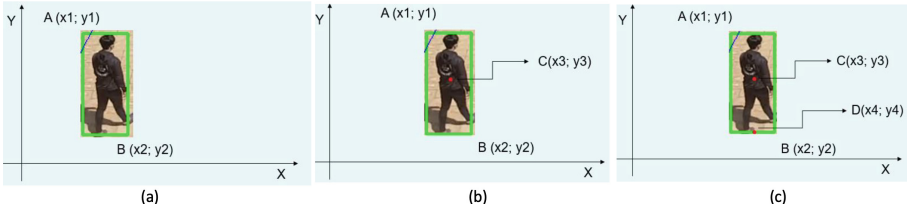### 3.3   Calculate Coordinates on a Detected Object

We used the pre-train model, Inceptionv2, to recognize and track people in the video to indicate whether they violated social distancing or not. After that process, we can get the bounding boxes corresponding to each object (person) in the region of interest (ROI). Then, to calculate the distance between 2 objects, we first have to determine the coordinates of each detected object. The coordinate calculation process consists of 2 steps. At first, we calculate the coordinates of the centroid of the bounding box with formulas Equations of 1 and 2. Then, After having the coordinates of the center of gravity of the bounding box, we go to find the coordinates of the midpoint of the bottom edge of the bounding box and use this point to calculate the distance between 2 objects on the birds'eye view as illustrated in Fig. 3 and calculated by Formulas of 3 and 4.

$$x3 = \frac{x1 + x2}{2} \tag{1}$$

$$y3 = \frac{y1 + y2}{2} \tag{2}$$

$$x4 = x3 \tag{3}$$

$$y4 = y3 + \frac{y2 + y1}{2} = y2 \tag{4}$$

Fig. 3. Calculate coordinates on a detected object: a) Present the original object in the coordinate axis, b) Calculate the coordinates of the center of gravity C(x3; y3) of the bounding box, c) Calculate the Midpoint coordinate D(x4; y4) of bottom edge bounding box

3.4   Calculate Distance Between 2 People

We build a matrix of 3×3 to perform coordinates transformation on 4 points with height and the width of the image[4]. To get the "birds'eyes view" from the top, we employ operations in OpenCV to calculate a perspective transform from four pairs of the corresponding points and generate a matrix of transformations of entities, including two parameters of rect, dst, where rect denotes the list of 4 points in the original image, and dst is the list of converted points. After getting the transformation matrix, we perform a perspective transformation to transform the image's perspective using the transformation matrix, along with the width and height of the image as input. Then, the operation returns the transformed matrix and image. Then there is a transformed matrix taken from the perspective transformation and a list of points to convert. Next, we take the output of the above calculation and start the process of passing the transformation calculation based on the entity transformation matrix: cv2.perspectiveTransform(list_points_to_detect, matrix) where *list_points_to_detect* is the result of the above calculation and *matrix* is the entity transformation matrix.

4   Experiments

4.1   Environmental Settings

We place the camera at 4.9m high from the ground at B as exhibited in Fig. 4. The camera was adjusted so that the length of the frame center to point A is 10.15 m (Fig. 5). In addition, the length from the center of the frame to point A can not affect the results of calculating the distance between objects (Fig. 7).

---

[4] https://pyimagesearch.com/2014/08/25/4-point-opencv-getperspective-transform-example/.

Fig. 4. camera's height

Fig. 5. Distance from camera to the center of bird's eye view.

We set up some procedures to calculate the minimum distance (two meters) between 2 people. First, we use a ruler with a length of two meters which is the minimum distance to prevent the risk of spreading COVID-19 as recommended by the Vietnamese Ministry of Health and numerous studies on Coronavirus disease. Then, we choose this as the minimum distance between the 2 received objects to start calculating the minimum distance. As illustrated in Fig. 6, the distance of 216 pixels is the distance calculated between 2 points from the bird's eye perspective.
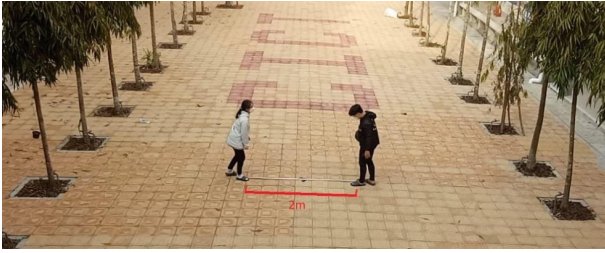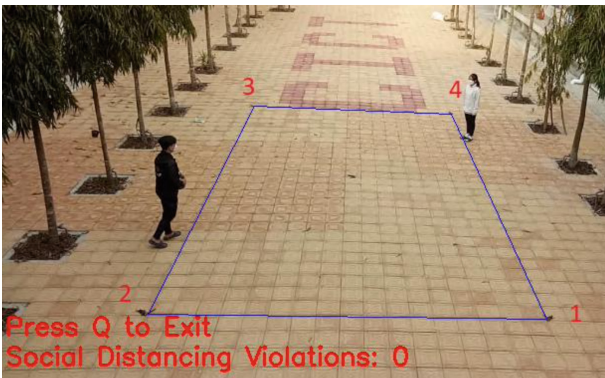
Fig. 6. We use a ruler to measure the distance in the pre-determined field

Fig. 7. Calculate the number of pixels of the minimum distance on the bird's eye view

Fig. 8. Four points for transformation and to indicate the pre-determined field.

In addition, we have deployed the process of converting pixels from the camera to birds'eye perspective and marked four points on the ground as four input points for perspective transition with perspective transition (Fig. 8). Here, choosing the order of each image angle is essential. If it is not selected correctly, it can reverse the image angle after conversion. The chessboard method can calibrate the camera to increase the model's accuracy. The method can be highly
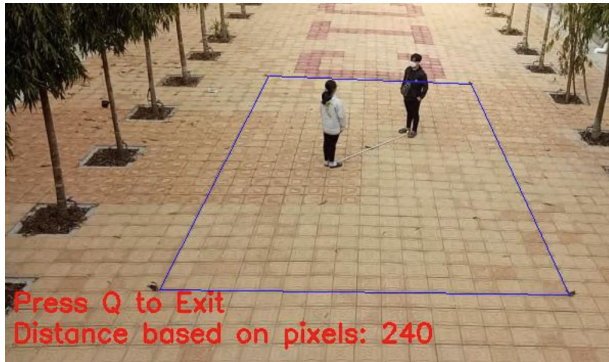
Fig. 9. The minimum distance to detect the violations on the birds'eye view.
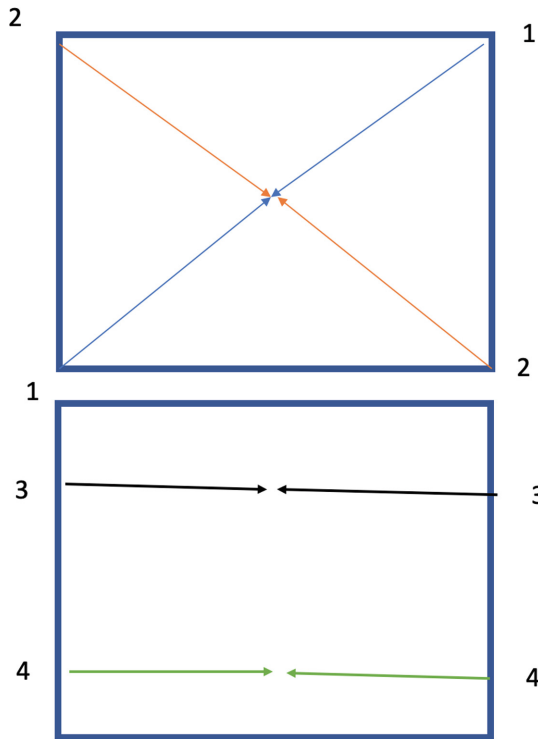

Fig. 10. Experimental participants' movement directions in Scenarios 1, 2, 3, and 4.

affected by the camera calibration and choosing the minimum distance directly from the birds'eye view. Here is an image when taking pixel parameters for a distance of two meters, as exhibited in Fig. 9. We obtain 240 pixels for a distance of two meters in the pre-determined field. Because the process of choosing the minimum distance is based on the distance of 2 points of the 2 rulers displayed on the video, the model's accuracy can depend on the accuracy of the camera calibration process during test cases. The accuracy of the calibration process reaches 0.23 (a smaller value can lead to a higher accuracy). We repeated the experiment more than 20 times for each scenario in the experiments. All of them were sent warnings to the users using the system. We illustrate some results in the following sections with experimental participants' movement directions as illustrated in Fig. 10.

4.2   Scenario 1

Fig. 11. A test case for Scenario 1.

One student moved from the top left corner and the other from the bottom right corner (Fig. 11). The two intersect at the center of the rectangle. We perform the experiments in full light conditions, with the participation of two people, one male, and one female, both wearing yellow shirts. The first object moves from the right side 1m from the bottom right corner. The second opponent moves 1m from the bottom right corner from the left side, and 2 objects move parallel to the bottom edge.

4.3   Scenario 2

One student moved from the top right corner while the other moved from the bottom left corner (Fig. 12). The two intersect at the center of the rectangle. When we test in full light conditions, with the participation of 2 subjects, 1 male and 1 female, wearing yellow shirts. The first enemy moves from the right

Fig. 12. A test case for Scenario 2.

side 1m from the top right corner, the second opponent moves from the left side 1m away from the top right corner, and 2 objects move parallel to the top edge.

4.4 Scenario 3

Fig. 13. A test case for Scenario 3.

One student moved from the midpoint of the upper edge, while the other moved from the midpoint of the lower edge (Fig. 13). Finally, the two intersect at the center of the rectangle. We perform the experiments in full light conditions, with the participation of two female and two male students. One pair wore yellow and black shirts (Fig. 14). The students moved from the 4 corners of the specified area, met, and stopped at the center of the specified area for about three seconds, and then the objects continued to move to the starting corner.
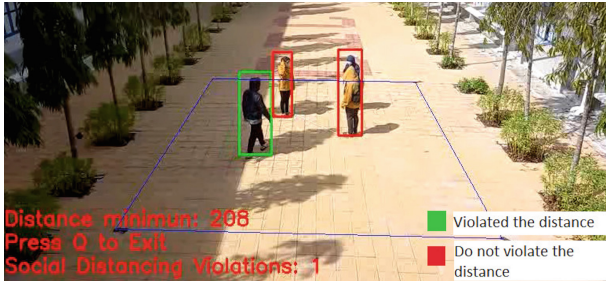
Fig. 14. Another test case for Scenario 3 with many people.

4.5   Scenario 4

Fig. 15. A test case for Scenario 4.

One student moved from the midpoint of the right side, and the other moved from the midpoint of the bottom left edge (Fig. 15). The two intersect at the center of the rectangle.

5   Conclusion

With the current epidemic situation being complicated with many outbreaks, even though we have a vaccine, keeping our distance still needs to be followed. The risk of transmission is rather low at two meters, so we should maintain a 2 m distance. In this study, we applied the pre-trained InceptionV2 model to recognize human objects in surveillance video, and then from the recognized human objects in each scene, we fine-tuned the camera. With methods of converting coordinates with a chessboard and bird's eye view, calculate distances

and provide warnings if violated. The method has been evaluated and tested with scenarios. However, the method's accuracy is based on the transition from the camera view to the birds'eye perspective. In addition, we convert measure units from meter to pixel to calculate distance. This can lead to dependency on the camera calibration process.

References

1. The Ministry of Health, Viatnamese: The ministry of health recommends "5k" to live safely with the epidemic (2020). https://covid19.gov.vn/bo-y-te-khuyen-cao-5k-chung-song-an-toan-voi-dich-benh-1717130215.htm

2. Bunn, S.: COVID-19 and social distancing: the 2 m advice (2020). https://post.parliament.uk/covid-19-and-social-distancing-the-2-metre-advice/

3. Payne, M.: What is the evidence to support the 2-m social distancing rule to reduce COVID-19 transmission? - a lay summary. https://www.healthsense-uk.org/publications/covid-19/204-covid-19-15.html

4. Chu, D.K., et al.: Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. **395** (10242), 1973–1987 (2020). https://doi.org/10.1016/s0140-6736(20)31142--9

5. Ahmed, I., Ahmad, M., Rodrigues, J.J.P.C., Jeon, G., Din, S.: A deep learning-based social distance monitoring framework for COVID-19. Sustain. Cities Soc. **65**(102571), 102571 (2021)

6. Goh, Y.M., Tian, J., Chian, E.Y.T.: Management of safe distancing on construction sites during COVID-19: a smart real-time monitoring system. Comput. Ind. Eng. **163**(107847), 107847 (2022)

7. Rosenfeld, D.L., Tomiyama, A.J.: Moral judgments of COVID-19 social distancing violations: the roles of perceived harm and impurity. Personal. Soc. Psychol. Bull. **48**(5), 766–781 (2021). https://doi.org/10.1177%2F01461672211025433

8. Elbishlawi, S., Abdelpakey, M.H., Shehata, M.S.: SocialNet: detecting social distancing violations in crowd scene on IoT devices. In: 2021 IEEE 7th World Forum on Internet of Things (WF-IoT). IEEE (2021). https://doi.org/10.1109

9. Acharjee, C., Deb, S.: YOLOv3 based real time social distance violation detection in public places. In: 2021 International Conference on Computational Performance Evaluation (ComPE). IEEE, December 2021. https://doi.org/10.1109

10. Bharti, V., Singh, S.: Social distancing violation detection using pre-trained object detection models. In: 2021 19th OITS International Conference on Information Technology (OCIT). IEEE, December 2021. https://doi.org/10.1109

11. Saponara, S., Elhanashi, A., Zheng, Q.: Developing a real-time social distancing detection system based on YOLOv4-tiny and bird-eye view for COVID-19. J. Real-Time Image Process. **19**(3), 551–563 (2022). https://doi.org/10.1007

12. Sriharsha, M., Jindam, S., Gandla, A., Allani, L.S.: Social distancing detector using deep learning. Int. J. Recent Technol. Eng. (IJRTE). **10**(5), 146–149 (2022). https://doi.org/10.35940

13. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48 T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48

14. Kumar, G., Shetty, S.: Application development for mask detection and social distancing violation detection using convolutional neural networks. In: Proceedings of the 23rd International Conference on Enterprise Information Systems. SCITEPRESS - Science and Technology Publications (2021). https://doi.org/10.5220/0010483107600767

15. Karaman, O., Alhudhaif, A., Polat, K.: Development of smart camera systems based on artificial intelligence network for social distance detection to fight against COVID-19. Appl. Soft Comput. **110**, 107610 (2021). https://doi.org/10.1016/j.asoc.2021.107610

16. Mercaldo, F., Martinelli, F., Santone, A.: A proposal to ensure social distancing with deep learning-based object detection. In: 2021 International Joint Conference on Neural Networks (IJCNN). IEEE (2021). https://doi.org/10.1109/ijcnn52387.2021.9534231

17. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning (2016)

18. Szegedy, C., et al.: Going deeper with convolutions (2014). https://arxiv.org/abs/1409.4842

19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). https://arxiv.org/abs/1409.1556

20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, pp. 1097–1105. NIPS 2012, Curran Associates Inc., Red Hook, NY, USA (2012)

21. Laureano, G.T., de Paiva, M.S.V., da Silva Soares, A., Coelho, C.J.: A topological approach for detection of chessboard patterns for camera calibration. In: Emerging Trends in Image Processing, Computer Vision and Pattern Recognition, pp. 517–531. Elsevier (2015). https://doi.org/10.1016

22. Bradski, G.: The OpenCV library. Dr. Dobb's J. Softw. Tools. **25**, 120–123 (2000)

23. Liu, Y., Liu, S., Cao, Y., Wang, Z.: Automatic chessboard corner detection method. IET Image Proc. **10**(1), 16–23 (2016). https://doi.org/10.1049/iet-ipr.2015.0126