Social Distancing Monitoring by Human Detection Through Bird's-Eye View Technique

Gona Rozhbayani[1], Amel Tuama[1] and Fadwa Al-Azzo[2]
[1]Kirkuk Technical Engineering College, Northern Technical University, Kirkuk 36001, Iraq
[2]Technical Engineering College Mosul, Northern Technical University, Mosul 41001, Iraq

Keywords: Human Detection, Social Distancing, Deep Learning, YOLOv5.

Abstract: The objective of this study is to offer a YOLOv5 deep learning-based system for social distance monitoring. The YOLOv5 model has been used to detect humans in real- time video frames, and to obtain information on the detected bounding box for the bird's eye view perspective technique. The pairwise distances of the identified bounding box centroid of people are calculated by utilizing euclidean distance. In addition, a threshold value has been set and applied as an approximation of social distance to pixels for determining social distance violations between people. The effectiveness of this proposed system is tested by experiments on different four video frames. The suggested system's performance showed a high level of efficiency in monitoring social distancing accurately up to 100%.

1 INTRODUCTION

Social distancing facilitates providing important information for a wide range of intelligent system applications, such as preventing cheating at the examination hall and maintaining privacy with others by not getting closer during transactions on Automated Teller Machines (ATMs). Additionally, it is a powerful method against the coronavirus disease in particular, which is declared a global pandemic by World Health Organization (WHO). In late 2019, Wuhan, China, received the first re- ports of it, which spreads due to contact with virus-infected people and also not following social distancing. In the fight against the COVID virus (Faiq et al., 2021), (Mustafa, 2021) social distancing has been proven to be a particularly successful strategy for controlling the disease's spread. People are encouraged to restrict their contact with one another for reducing the chances of the virus being transmit- ted by direct or close personal contact. Be advised that papers in a technically unsuitable form will be returned for retyping. After returned the manuscript must be appropriately modified. The general population is not accustomed to enclosing themselves in a protective bubble. The perceptive capabilities of humans can be improved and helped by an automated warning system. As it is well said, "prevention is better than cure", and the WHO has recommended many precautions to reduce coronavirus transmission. In the current environment, social distancing and separation (Ecdc & UwKr, 2020) have been shown to be one of the most effective spread preventers. Therefore, to follow social distancing, everyone should keep a space of at least 6 feet (2 meters) between them, ac- cording to WHO's standard prescriptions. This is a well-known method of breaking the chain of infection. For that reason, social distance has become the norm in all afflicted countries. Artificial intelligence (AI) with Deep Learning (DL) methods has shown hopeful outcomes in several daily life problems (Awad et al., 2022), (Hasan et al., 2020) where deep learning is a feature of artificial intelligence that replicates how the human brain processes data and looks for things. The AI applications can be used in many places such as offices, organizations, malls, educational institutes, banks, etc. Besides DL handles the information faster and more effectively than humans in monitoring social distancing. Monitoring social distancing is difficult to keep in real-time circumstances. There are two ways to do it: manually and automatically. To guarantee that everyone is properly following the social separation guidelines, the manual approach requires a huge number of physical eyes. This is an impossible task because the observer cannot keep their eyes open 24 hours a day for constant

Rozhbayani, G., Tuama, A. and Al-Azzo, F.
Social Distancing Monitoring by Human Detection Through Bird's-Eye View Technique.
DOI: 10.5220/0012373900003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 4: VISAPP*, pages 306-313
ISBN: 978-989-758-679-8; ISSN: 2184-4321

monitoring, instead of automated surveillance systems (Singh & Kushwaha, 2016) can be used that replace many physical eyes with closed-circuit television cameras (CCTV) so that the system provides alerts if specific distance not maintained. The proposed work aims to reduce the COVID virus spread by detecting people who violate social distancing rules using the YOLOv5 model with a real-time system. Bird's-eye view perspective measurement is added to the detection system, consequently, accurate detection results are produced according to the image's perspective. The remaining of this paper is arranged as follows. The highlights of the earlier studies and related works are given in Section 2. In Section 3, we present the YOLOv5 method and the detail of the proposed system. In section 4 we describe and analyse the experimental results. Section 5 presents the conclusion of this paper.

2 RELATED WORK

Social distancing has been the subject of several studies using a variety of methodologies. In order to automate people detection in crowded areas, both inside and outdoors, using standard CCTV cameras, (Rezaei & Azarmi, 2020) suggested the framework that employs the YOLOv4 based Deep Neural Network (DNN) model. For people detection and social distance monitoring, they combined the DNN model with an adjusted inverse perspective mapping (IPM) method and SORT tracking algorithm. As well, a framework using the YOLOv4 model for real-time object identification was suggested by (Rahim et al., 2021). In their YOLOv4 model framework, they also presented the social distance measurement technique to specify the risk factor based on the computed distance and safety distance violations. To capture the video sequence un- der diverse lighting circumstances, one motionless time of flight (ToF) camera was added to this model. An overhead viewpoint is used by (Ahmed et al., 2021) to detect individuals and monitor their social distance using a deep learning platform based on the YOLOv3 model. Using an overhead dataset, the detection approach connects a pre-trained model to an additional trained layer. The detection model uses bounding box data to detect humans. Human pairwise distances are measured by applying the euclidean distance for the bounding box centroid values to detect violations of social distance. Additionally, (Saponara et al., 2021) research indicates utilizing thermal pictures to classify people according to their social distancing using artificial intelligence. A unique strategy of deep learning detection is created for detecting and tracking individuals in both indoor and outdoor environments by implementing the YOLOv2 model. The offered approach is utilized to build a full AI system for people tracking, social distancing classification, and monitoring of body temperature using images captured by thermal cameras.

3 METHODOLOGIES

3.1 YOLOv5 Architecture

The most fundamental version of the object detection strategy is known as You Look Only Once (YOLO)v5 (Wu et al., 2021). The image's location and object class are determined by applying a Convolution Neural Network (CNN) architecture to it. The three essential parts of the YOLOv5 architecture are the backbone, the neck, and the head.

3.1.1 The Backbone

The backbone is mostly used to extract important features from the supplied input image. Convolution, Convolution Three (C3), and Spatial Pyramid Pooling (SPP) are some of the building components that make up the YOLOv5 backbone structure. The smallest component of the YOLO network structure is the Convolution structure as demonstrated in Fig. 1, which processes the received image with a resolution of 640×640×3. It first becomes a 320×320×12 feature map via the slicing operation, and then a convolution procedure utilizing 32 convolution kernels results in a 320×320×32 feature map. The convolution block makes use of a specific structural diagram composed of a convolution layer, a batch normalization (BN) layer, and an activation function layer called a sigmoid linear unit (SiLU). CSP Bottleneck and three convolutions, is present in the C3 block as illustrated in Fig. 2. The SPPF uses three separate 5*5 maximum poolings, where the input is sent through each one in serial to speed up calculation. The architecture of the YOLOv5s model is shown in Fig. 3.

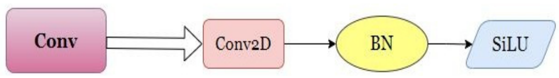


Figure 1: Structure of Convolutional Block.

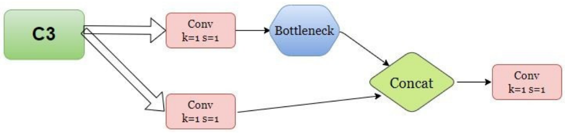


Figure 2: Structure of C3 Block.

3.1.2 The Neck

This part of the model is used to create feature pyramids. Using feature pyramids improves the model's ability to generalize on object scaling and perform well with unknown data. It facilitates the identification of the same object at various sizes and scales. Other models employ a variety of feature pyramid approaches, including FPN, PANet, etc. As the image progresses through the different layers of the neural network levels, the features complication raises, and the spatial resolution of the image decreases concurrently. As a result, the high-level features are unable to precisely identify the pixel-level masks.

3.1.3 The Head

The head is mostly necessary to complete the final phase of prediction. It is made up of three convolution blocks, and the bounding box, confidence score, and label estimates are all represented as vectors. Boxes with low scores are discarded to finish the output processing.

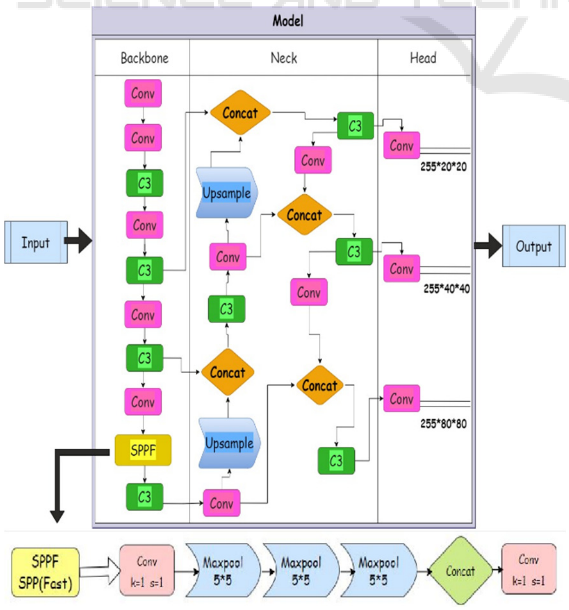
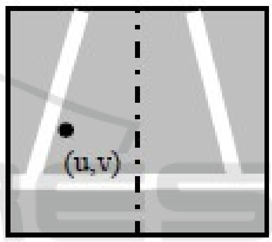


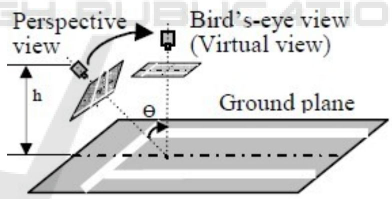
Figure 3: YOLOv5s Model Structure.

3.2 Bird's-Eye View

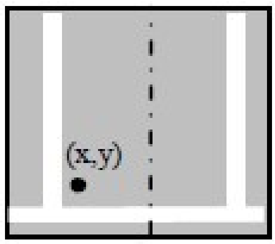
The proposed work in this paper utilizes the Bird's-eye view technique (M.Venkatesh, 2012). In order to get better in- sights into the required information, the changing perspective of a given frame in the video was carried out. For example, if the camera is mounted at a 45 angle to the ground and rotation of the image by 45 angles is recommended. The detection procedure will be handled as the captured images were taken somewhere around a 90-degree angle from the above view. After that, the top view of the given image is obtained. The technique is applied by defining four coordinate points in the front view image (Luo et al., 2010), then the perspective of the image is adjusted to the needed view as shown in Fig. 4. As a result, the detected humans were marked in a circle of their initial form perspective. Accordingly, more precise detection results have been obtained utilizing this bird's-eye view standpoint technique.



a) Image of perspective view.



b) Perspective transformation.



c) Image of bird's eye view.

Figure 4: Illustration of perspective transformation.

3.3 Social Distancing Measuring

Identifying coordinates works with the bounding box of a detected person. The method helps in determining if detected people in each frame are in safe distancing or not. The distance between the centers of each detected person was calculated, therefore identifying the centroid value of different object bounding boxes used for measuring centroid as shown in Eq. (1).

$$C(C_X, C_Y) = \frac{X_L + X_R}{2}, \frac{X_T + X_B}{2} \quad (1)$$

Where bounding box coordinates include XL (Left), YT (Top), XR (Right), and YB (Bottom).

The euclidean distance formula is employed to calculate the spacing between the center of the bounding box as shown in Eq.2. After that the measured distance has been compared with a predefined threshold value. A threshold value has been set and applied as an approximation of social distance to pixels to determine social distance violations between people. If the distance is greater than the threshold value, it indicates that the detected people are safe. While in case the distance is less than the threshold value, that means detected people are not following the social distancing rule.

$$\text{Euclidean Distance } (D) = \sqrt{(C_{X2} - C_{X1})^2 + (C_{Y2} - C_{Y1})^2} \quad (2)$$

Where (CX1, CX2) and (CY1, CY2) are centroid values of two persons.

3.4 Dataset

Two different custom datasets are used to evaluate the proposed system, as demonstrated in Fig. 5, and 6. The Open Images Datasets (Kuznetsova et al., 2018) was the first dataset that utilized and composed of 1100 images, while the second dataset was gathered from Kirkuk Technical Engineering College (KTEC) and contains 3354 images. The datasets include several difficult recording situations, including (indoor, and outdoor, various outfits, low resolution, both genders, etc). The dataset pre-processing is also obtained to resize each image to 640*640 in order to enhance the training efficiency model. To evaluate the performance of the candidate model, the datasets were then split into a training set and validation set, each comprising 80% and 20% of the total datasets. Using the Roboflow platform the human image was manually tagged to produce the training dataset. which is a form of computer vision annotation tool and a computer vision developer framework (CVAT) due to the necessity of knowing the object's (person's) true identity. As seen in the Fig.7, it is crucial to create a bounding box (BBox) around just the per- son and ignore the other items in the same image. As a result, the suggested system recognizes that the "person" inside the box belonged to the person class. Each image's annotations are saved by Roboflow in ".txt" text file format. The format is shown as "objectclass-ID" "X center" "Y center" "Box width" and "Box height", For every human in the image, one row of a BBox annotation is present in each text file. Then, a labels folder is created and contains all ".txt" files.

Figure 5: The OID Dataset Samples.

Figure 6: The KTEC Dataset Samples.

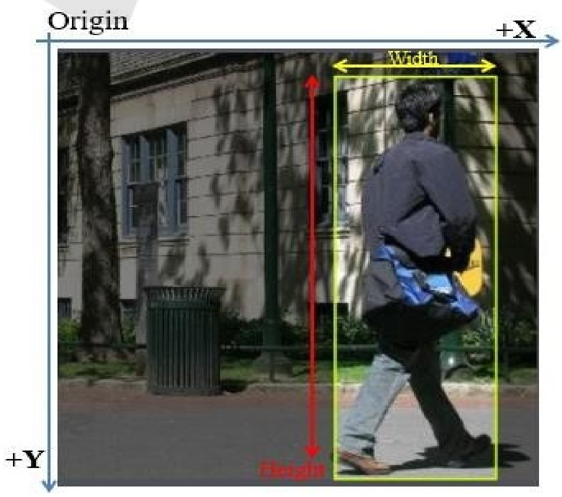


Figure 7: Person class sample with a Bounding Box surrounding.

3.5 Proposed System

The operational steps of the suggested system in this work shown in Fig. 8 aim to detect people who violate or inviolate the social distancing rule based on deep learning techniques Words like “is”, “or”, “then”, etc should not be capitalized unless it is the first word of the sub subsection title.

Measurement of perspective is the first step to take which is called the Bird’s-eye view technique, a quadrangular area of the image is selected that corresponds to what would be similar to a rectangle in reality. Then, the perspective transformation occurred to the quadrangular area, it appeared as being observed from the above view, much like a bird, as shown in Fig.9.

The YOLOv5 model has been used in the proposed system to detect a human in real-time. The model is taught to provide the required results as part of the training process by locating acluster of weights in the network that are appropriate for dealing with certain problems. In order to identify problems with human detection, the YOLOv5s model was fine-tuned for two custom datasets. After that, coordinates identification is appliedto gain the centroid of each detected person from the bounding box. The euclidean distance formula is used to calculate the distance between every two persons with the threshold value. The threshold value setting changed from image to image and was regulated by the camera’s position and its proximity to peo ple. cv2.line( ) function is a function in the OpenCV-Python library, which is used to draw a line on the image for getting the required length (threshold value) through trial and error as illustrated in Fig. 10. The actual distance is converted to pixel distance to decide on the right threshold value for each tested case. However, the pixel numbers for the required distance of tested cases are taken by USB webcam 034 camera model resolution with the raspberry pi 4 model B, pixel numbers changealong with changes in camera resolution. Eventually, the system will decide on safe or risky social distance based on the number of humans and the measured distance between them bycircles. If the social distancing rule is violated and people are too close to each other, the circle will change from green to redcircle.

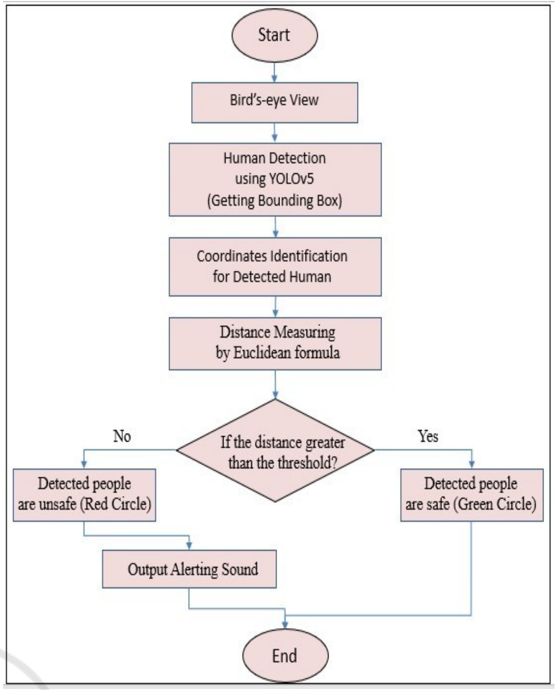


Figure 8: Flowchart of the system.

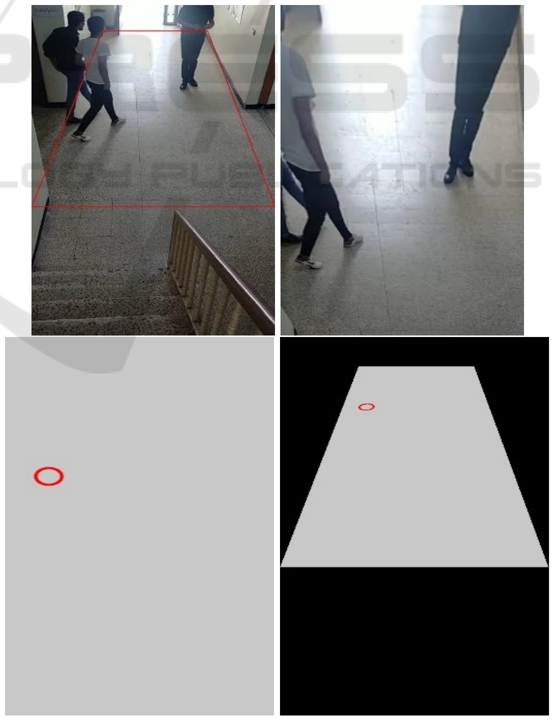
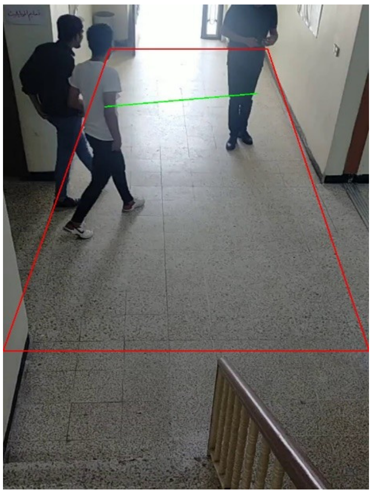


Figure 9: Quadrangular area with bird’s-eye view perspective.



a) First Example.

b) Second Example.

Figure 10: Defining Threshold Value.

4 EXPERIMENTAL RESULTS

The proposed system efficiency was tested and analyzed based on actual real-time video frames in diverse situations and from various camera positions, along with the distance between the camera and objects. The threshold value of each tested case is set by the trial-and-error method to find the adequate threshold value. The obtained results are computed according to the proportion of two sorts of detections, human detection by utilizing the YOLOv5 model and social distancing detection. How- ever, accurate object detection outcome results are essential for a social distancing system. The testing success average on this system is estimated by applying the formula as shown in Eq. (3)

$$R = \frac{SD}{ND} * 100\% \tag{3}$$

Where R is the proportion of experiment successful results, SD is the count of successful detections, ND is the count of noticed data.

The first case video is captured at a distance of around 3 to 5 meters far away from people, as shown in Fig. 10. As well as, the position of the camera was parallel to humans. The ratio of successful detection by the YOLOv5 model accomplished 100%. Afterward, the social distancing outcome reached 100% likewise. If the distance between two people is greater than the threshold value which equals (200 pixels), thus they are at a safe distance from each other. As a result, the system will be denoted by green circles as shown in Fig. 11.

Figure 11: First Tested Case.

The second case video is captured in an outdoor environment, approximately 7 to 10 meters between the camera and objects. The camera is positioned such that it was parallel to humans. The achieved accuracy of human detection and social distancing results were both 100%. As shown in Fig. 12 two people are too close to each other which leads to violating the social distancing rule according to the distance between them being less than the threshold value (160 pixels). As a result, the system marked the unfollowed people for the distancing rule with red circles.

Figure 12: Second Tested Case.

The third video is taken in an indoor environment at a distance of around 8 to 10 meters far away from objects, the camera was placed at a higher level compared to humans. At this substantial distance, the human detection accuracy accomplished 100% for the performance of the video test, indicating that Yolo-v5 can identify many persons from a considerable distance shown in Fig. 13. Among the people in the video, one of them is far enough to not violate the distancing rule, then the system marked the person in a green circle. However, other people were close to each other their distance was less than the threshold value (160 pixels), thus they indicated by red circles.

Figure 13: Third Tested Case.

The fourth case is held inside a shopping center within many people. The distance of the camera to the people was about 10 to 15 meters shown in Fig. 14, the camera laid approximately in the range of 70 to 75-degree angle. Therefore, the people appeared too small due to the farness of objects from the camera and the threshold value set (120 pixels). Thus, the obtained human detection result by using the YOLOv5 model gained 68.1%. While the accuracy of the social distancing detection achieved 100% since all detected humans were discovered right whether they had violated or not.

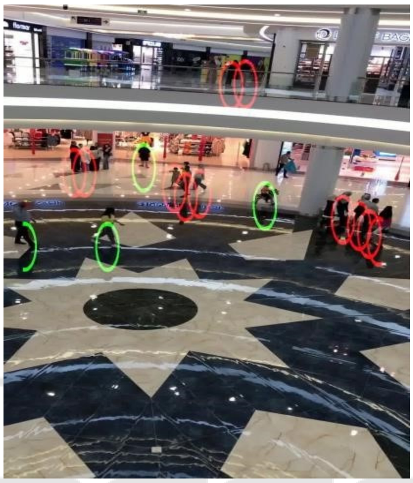


Figure 14: Fourth Tested Case.

The comparison of all the obtained results on the tested cases that have been carried out are summarized in Table 1 as follows:

Table 1.

Test Cases	Threshold Value	People Distance to Camera	Human Detection Accuracy	Social Distancing Detection Accuracy	Detected People in the Frame	Safe People	Unsafe People
First case	200	3-5m	100%	100%	2	2	0
Second case	160	7-10m	100%	100%	4	2	2
Third case	160	8-10m	100%	100%	4	1	2
Fourth case	120	10-15m	68%	100%	15	4	11

5 CONCLUSIONS AND FUTURE WORK

This paper introduced a system to monitor social distancing in real-time videos based on a deep learning technique, the system utilized the YOLOv5 model for human detection. In addition, the bird's eye view perspective assisted the Euclidean formula to obtain more accurate results in the calculating distance between two persons. According to the outcomes of the testing, the distance between the camera and the human has an impact on the accuracy of the system. The accuracy of the system performance was 100% by utilizing the camera to human distance from 3 to 10 meters, While, if the people were too far away from the camera, they appear too tiny. Consequently, the human detection accuracy decreased at a distance of 10 meters or more, because they could not be detected by the model. As a result, the social distancing system performance declines as well.

In future work, simultaneous human detection and tracking with multiple cameras could be added to this system. For this reason, in order to undertake pedestrian matching, features of detected bounding boxes in successive frames would be extracted and compared with data from other cameras.

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. https://doi.org/10.1016/j.scs.2020.102571

Awad, S. R., Sharef, B. T., Saiih, A. M., & Malallah, F. L. (2022). DEEP LEARNING-BASED IRAQI BANKNOTES CLASSIFICATION SYSTEM FOR BLIND PEOPLE. Eastern-European Journal of Enterprise Technologies, 1(2–115), 31–38. https://doi.org/10.15587/1729-4061.2022.248642

Ecdc, & UwKr. (2020). S. R. Awad, B. T. Sharef, A. M. Salih, F. L. Malallah, Deep learning-based iraqi banknotes classification system for blind people, Eastern-European Journal of Enterprise Technologies 1 (2022) 115.

Faiq, T. N., Ghareeb, O. A., & Fadhel, M. F. (2021). The relationship between blood groups, lifestyle and infection with covid-19 in nineveh governorate-iraq, (Vol. 25). http://annalsofrscb.ro

Hasan, S., Mahmood, S., Rafi, R., Al-Nima, O., Hasan, S. Q., & Esmail, S. (2020). Exploiting the Deep Learning with Fingerphotos to Recognize People Training Scientific Research View project Diploma Artificial Intelligence View project Exploiting the Deep Learning with Fingerphotos to Recognize People. International Journal of Advanced Science and Technology, 29(7), 13035–13046. https://www.researchgate.net/publication/345236222

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., & Ferrari, V. (2018). The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. https://doi.org/10.1007/s11263-020-01316-z

Luo, L. B., Koh, I. S., Min, K. Y., Wang, J., & Chong, J. W. (2010). Low-cost implementation of bird's-eye view system for camera-on-vehicle. ICCE 2010 - 2010 Digest of Technical Papers International Conference on Consumer Electronics, 311–312. https://doi.org/10.1109/ICCE.2010.5418845

Mustafa, K. N. (2021). The Relationship between Blood groups, Lifestyle and infection with Covid-19 in Nineveh Governorate-Iraq (Vol. 25). http://annalsofrscb.ro

M.Venkatesh, P. V. (2012). 3 ROTATION AND SCALING OF AN IMAGE. http://www.ijser.org

Rahim, A., Maqbool, A., & Rana, T. (2021). Monitoring social distancing under various low light conditions with deep learning and a single motionless time of flight camera. PLoS ONE, 16(2 February). https://doi.org/10.1371/journal.pone.0247440

Rezaei, M., & Azarmi, M. (2020). Deepsocial: Social distancing monitoring and infection risk assessment in covid-19 pandemic. Applied Sciences (Switzerland), 10(21), 1–29. https://doi.org/10.3390/app10217514

Saponara, S., Elhanashi, A., & Gagliardi, A. (2021). Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19. Journal of Real-Time Image Processing, 18(6), 1937–1947. https://doi.org/10.1007/s11554-021-01070-6

Singh, D. K., & Kushwaha, D. S. (2016). Tracking movements of humans in a real-time surveillance scene. Advances in Intelligent Systems and Computing, 437, 491–500. https://doi.org/10.1007/978-981-10-0451-3_45

Wu, T. H., Wang, T. W., & Liu, Y. Q. (2021). Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network. 2021 3rd World Symposium on Artificial Intelligence, WSAI 2021, 24–28. https://doi.org/10.1109/WSAI51899.2021.9486316.